

ABSTRACT OF THE DISCLOSURE

A system and method for identifying a small group of compounds representative of a larger set of compounds is disclosed. The system obtains one or more descriptors, determines the median value for the values of each descriptor for a set of compounds, partitions the set of compounds into a plurality of partitions using each median value for the set of compounds, and selects compounds from each of the partitions to form a subgroup representative of the set of compounds. A system and method for virtual compound screening is also disclosed. The system recursively partitions a set of compounds based on descriptor median values where the partitions which have at least two bait compounds are recombined and repartitioned until a desired number of compounds remain in the partition.